Homology Probing: Identification of cDNA Clones Encoding Members of the Protein-Serine Kinase Family

Steven K. Hanks

Proceedings of the National Academy of Sciences of the United States of America, Volume 84, Issue 2 (Jan. 15, 1987), 388-392.

Stable URL:
http://links.jstor.org/sici?sici=0027-8424%2819870115%2984%3A2%3C388%3AHPIOCC%3E2.0.CO%3B2-4

Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 388–392, January 1987
Biochemistry

Homology probing: Identification of cDNA clones encoding members of the protein-serine kinase family

(oligonucleotide probes/HeLa cells/phosphorylase kinase/*CDC2* and *CDC28* genes)

STEVEN K. HANKS

Molecular Biology Laboratory, Salk Institute for Biological Studies, P.O. Box 85800, San Diego, CA 92138

Communicated by Robert W. Holley, October 3, 1986 (received for review August 25, 1986)

ABSTRACT Mixed oligonucleotide probes were used to screen a HeLa cDNA library for clones encoding amino acid contiguities whose conservation is characteristic of the protein-serine kinase family. Eighty thousand clones were screened, from which 19 were identified as showing strong hybridization to two distinct probes. Four clones were chosen for characterization by partial DNA sequence analysis and 3 of these were found to encode amino acid sequences typical of protein-serine kinases. One deduced amino acid sequence shares 72% identity with rabbit skeletal muscle phosphorylase kinase γ-subunit, while another is closely related to the yeast protein-serine kinases CDC2 in *Schizosaccharomyces pombe* and CDC28 in *Saccharomyces cerevisiae*. This screening approach should have applications in the identification of clones encoding previously unknown or poorly characterized members of other protein families.

Protein phosphorylation events are of central importance in the response of cells to various internal and external signals (reviewed in refs. 1 and 2). The number of unique protein kinases and protein kinase activities that have been described now exceeds 50. These can broadly be classified into two subfamilies based on substrate specificity; protein-serine kinases phosphorylate serine and to a lesser extent threonine residues, while protein-tyrosine kinases specifically modify tyrosine. Characterization commonly involves protein-purification methodology coupled with a suitable *in vitro* assay by which substrate-specific phosphate additions can be monitored. In some instances, investigators have succeeded in purifying a particular protein kinase (3–6) or protein kinase-derived peptide (7–10) to a degree and quantity sufficient for amino acid analysis. However, most protein kinase activities have not been highly purified and characterization is limited to biochemical studies. Another route to protein kinase identification and characterization stems from the study of oncogenic retroviruses. Many retroviral transforming proteins and their cellular counterparts have been shown to possess protein kinase activity (reviewed in refs. 11 and 12). Amino acid sequences for this group have been deduced from DNA coding sequences, circumventing the often laborious protein purification procedures.

The availability of amino acid sequence data for a number of protein kinases leads to an alternative approach to the identification and subsequent characterization of additional members of this family of enzymes. Alignment of the sequences for maximum homology reveals several short stretches within the catalytic domain where amino acids are highly conserved throughout the protein kinase family (5, 12, 13). Furthermore, some residues within these stretches appear to distinguish between subfamilies, being specific for either protein-serine or protein-tyrosine kinases. DNA sequences encoding these short stretches of homology can serve as specific hybridization targets for synthetic oligonucleotide probes and, thereby, DNA clones encoding members of a particular subfamily can be identified from a large library of sequences. I have tested the validity of this approach by screening a HeLa cDNA library with probes designed to recognize clones encoding protein-serine kinases.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

METHODS

cDNA Library Construction. HeLa cells, obtained from the American Type Culture Collection, were grown in Dulbecco's modified Eagle's medium supplemented with 10% calf serum. Total RNA was isolated from an exponentially growing cell population using the guanidinium thiocyanate method (14) and poly(A)$^+$ RNA was selected by two rounds of oligo(dT) chromatography (15). The poly(A)$^+$ RNA was used to construct a cDNA library as described by Okayama and Berg (16). The cDNA constructs were transformed into *Escherichia coli* strain DH-5 using the high efficiency method of Hanahan (17).

Library Screening with Oligonucleotide Probe Mixtures. Oligodeoxynucleotide mixtures were synthesized by the phosphoramidite method (18) on a Systec model 1450A synthesizer. The cDNA library was initially screened by colony hybridization (19) on nitrocellulose filters using oligonucleotide mixtures labeled with ^{32}P by polynucleotide kinase (20). Hybridizations were carried out at 37°C in 5× SSPE/5× Denhardt's solution (0.74 M NaCl/50 mM $NaH_2PO_4 \cdot H_2O$/5 mM $Na_2EDTA \cdot 2H_2O$/0.1% Ficoll/0.1% polyvinylpyrrolidone/0.1% bovine serum albumin) supplemented with hydrolyzed yeast RNA (100 μg/ml) (21) and 0.05% $Na_4P_2O_7 \cdot 10H_2O$ as additional blocking agents. After hybridization, filters were washed in 3 M $(CH_3)_4NCl$/50 mM Tris·HCl, pH 8.0/2 mM $Na_2EDTA \cdot 2H_2O$/0.1% $NaDodSO_4$ according to Wood *et al.* (22). The wash temperature was set at 52°C to select for long contiguous matches. Autoradiography was for 16–24 hr at room temperature.

For secondary screening, plasmids were first purified from positive colonies by a small-scale alkaline lysis procedure (23). Plasmids were digested with *Pst* I, then DNA fragments were separated by agarose gel electrophoresis and transferred to nitrocellulose (24). Hybridizations to oligonucleotide probes and washing solutions were as described for the colony hybridization procedure. Sequential washings of increasing stringency were performed by increasing the wash temperature from 42°C to 47°C to 52°C, between which the blots were processed for autoradiography. To allow rehybridization of the filters, probes were removed by denaturation (25).

DNA Sequence Analysis. DNA sequence analysis was by the dideoxy-chain termination method (26) following subcloning into the *Pst* I site of M13mp18 (27). The orientations of inserts in the single-stranded templates were determined by dot hybridization (28) using ^{32}P-labeled oligonucleotide

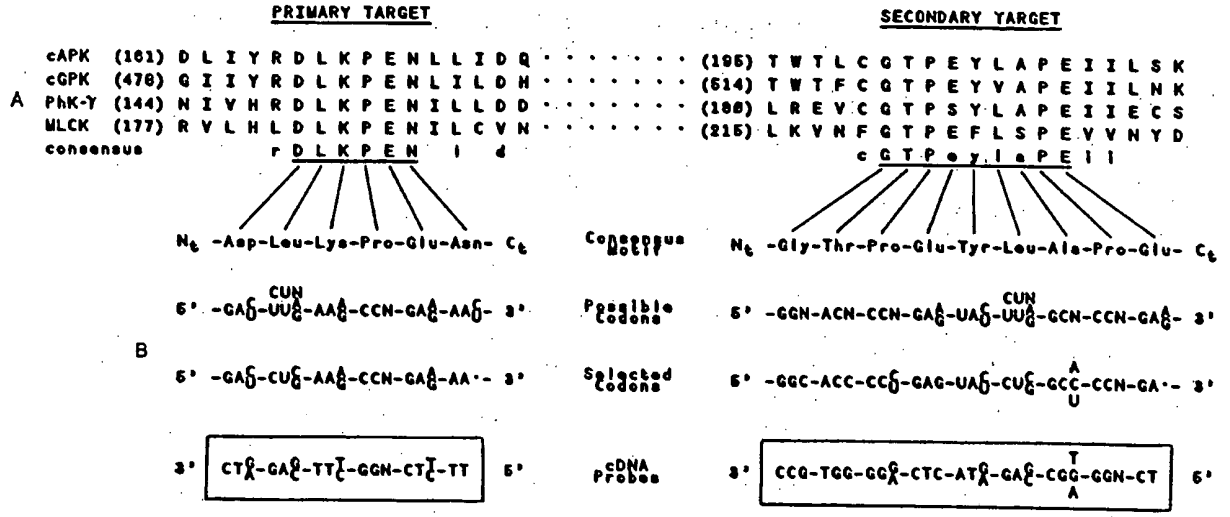

FIG. 1. (*A*) Alignment of amino acid sequences from the catalytic domains of four mammalian protein-serine kinases reveals two separate highly conserved segments. Sequences shown are as follows: cAPK, cAMP-dependent protein kinase from bovine cardiac muscle (3); cGPK, cGMP-dependent protein kinase from bovine lung (5); PhK-γ, phosphorylase kinase γ subunit from rabbit skeletal muscle (4); MLCK, myosin light-chain kinase from rabbit skeletal muscle (6). The single-letter amino acid code is used. Only the conserved stretches and surrounding residues are shown. Numbers in parentheses indicate amino acid position from amino termini. For MLCK, where the amino-terminal portion of the protein has not been determined, numbering initiates at the amino-terminal end of the reported sequence. Residues conserved in each of the four sequences are shown in the consensus line as uppercase letters and conservation in three of the four sequences is indicated with lowercase letters. Consensus sequences chosen for probe targeting are underlined. (*B*) Oligonucleotide probe design. The primary probe is a mixture of 64 17-mers complementary to RNA sequences encoding the conserved sequence: Asp-Leu-Lys-Pro-Glu-Asn (D L K P E N). All possible codons are represented except for the leucine position, where two of the six possible codons were selected. The two selected leucine codons are used at a combined frequency of 0.68 in human protein-coding sequences (29). More extensive codon selection was used in designing the secondary probe, a mixture of 96 26-mers complementary to RNA sequences encoding the consensus sequence Gly-Thr-Pro-Glu-Tyr-Leu-Ala-Pro-Glu (G T P E Y L A P E). The minimum pairing frequency between the most complementary oligonucleotide in the secondary mixture and a target sequence is 0.73 (19 of 26 bases paired), while the predicted best-pairing frequency, based on codon usage (29), is 0.91. Oligodeoxynucleotide mixtures used as probes are enclosed in boxes.

probes and hybridization conditions as described above. Sequencing from templates with inserts in each orientation enabled sequence determination for the entire insert. Performing the sequencing reactions at 50°C aided in reading through oligo(dG) tails present in some of the inserts.

RESULTS

Design of oligonucleotide probes was based on amino acid sequence alignment (5, 12) of catalytic domains from four mammalian protein-serine kinases (Fig. 1*A*). Two highly conserved segments were selected for oligonucleotide probe targeting. The primary target was chosen as nucleotides encoding the amino acid sequence: Asp-Leu-Lys-Pro-Glu-Asn (single letter code, D L K P E N). This sequence, beginning at position 166 in cAMP-dependent protein kinase (3), is also contained in each of three other reported protein kinase sequences: cGMP-dependent protein kinase (5), phosphorylase kinase γ-subunit (4), and myosin light-chain kinase (6). In the protein-tyrosine kinases, the two sequences Asp-Leu-Arg-Ala-Ala-Asn (D L R A A N) and Asp-Leu-Ala-Ala-Arg-Asn (D L A A R N) are most characteristic of this conserved region (12), and this subfamily therefore will not be recognized by probes directed toward the D L K P E N homology.

The secondary target sequence encodes the consensus motif Gly-Thr-Pro-Glu-Tyr-Leu-Ala-Pro-Glu (G T P E Y L A P E). This region of homology lies 29 residues COOH-terminal to the D L K P E N sequence in cAMP-dependent protein kinase and shows a small level of divergence among the other protein-serine kinases (Fig. 1*A*). Among the protein-tyrosine kinases, this region has diverged considerably (12) and is typified by the sequence Gly-Ala-Lys-Phe-Ile-Lys-Trp-Thr-Ala-Pro-Glu (G A K F P I K W T A P E) found in the transforming protein of Rous sarcoma virus (30).

To ensure high match frequencies, mixtures of oligonucleotides representing many of the codon possibilities were used as probes (Fig. 1*B*). Codon selection was based on utilization frequencies in human protein-coding sequences (29). Specificities of probes were tested by conducting a computer-assisted homology search against the National Institutes of Health GenBank library.*

Initially, the D L K P E N probe was used to screen a HeLa cell cDNA library by colony hybridization. Following a high-stringency wash, a positive signal was detected in 89 of ≈80,000 colonies screened. Plasmids from each positive clone were purified to facilitate secondary screening. Each plasmid was cleaved with *Pst* I and DNA fragments were separated by agarose gel electrophoresis (Fig. 2 *Upper*). The fragments were then transferred to nitrocellulose filters and hybridized to the secondary G T P E Y L A P E probe (Fig. 2 *Middle*). Of the 89 positive clones identified with the D L K P E N probe, 19 also showed significant hybridization to the G T P E Y L A P E probe.

To identify small fragments recognized by both probes, the G T P E Y L A P E probe was removed by denaturation and the blots were rehybridized with the D L K P E N probe (Fig. 2 *Lower*). Several small *Pst* I fragments were detected that hybridized strongly to both. Four such fragments ranging in size from 300 to 450 base pairs (bp) were chosen for preliminary DNA sequence analysis. These fragments represented four individual clones: PSK-C3 (*putative protein-serine kinase*; filter *C*, colony *3*), PSK-H1, PSK-J3, and

*National Institutes of Health (1985) Genetic Sequence Databank: GenBank (Research Systems Div., Bolt, Beranek, and Newman, Inc., 10 Moulton St., Cambridge, MA 02238), Tape Release 38.

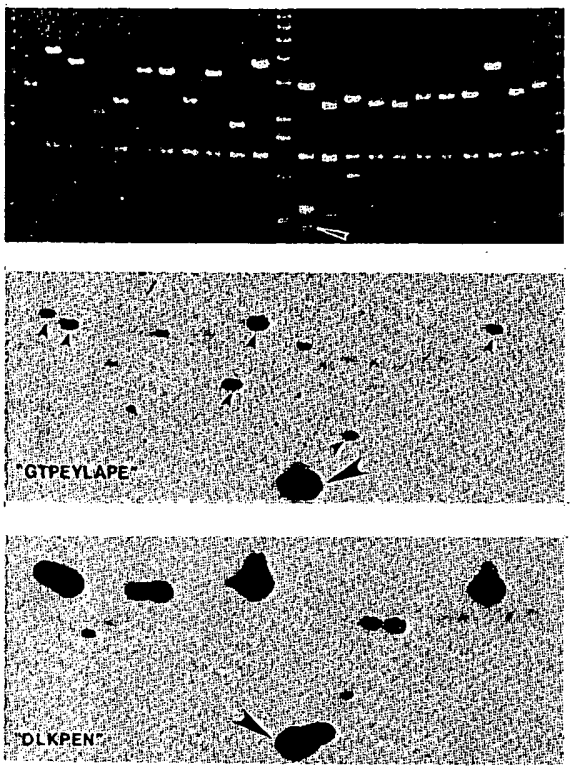

FIG. 2. Secondary screening of positive clones. Results for only 22 of the 89 clones are shown. (*Upper*) Ethidium bromide-stained plasmid DNA fragments after digestion with *Pst* I and separation by agarose gel electrophoresis. Two *Pst* I cutting sites are present in the vector, giving rise to the 1.3-kbp band present in all lanes and a larger fragment containing both vector and cDNA insert sequences. Additional bands arise from cutting sites within the insert. The central and two end lanes contain molecular size standards (Bethesda Research Laboratories' 1-kb ladder fragments shown are, from major band at bottom, 0.51, 1.02, 1.64, 2.04, 3.05, 4.07, 5.09, 6.11, 7.13, and 8.14 kbp). (*Middle*) Pattern of hybridization to the G T P E Y L A P E probe mixture. Several fragments give strong hybridization signals (arrows). (*Lower*) The same blot rehybridized to the D L K P E N primary probe mixture. The 450-bp fragment indicated by the open arrow (*Upper*) shows very strong hybridization to both probes (large arrows). This fragment is part of an ≈3.2-kbp insert from the plasmid designated PSK-H1.

PSK-K5. The fragments were subcloned into M13 and single-stranded templates were prepared from several plaques containing each PSK fragment. As a final screening procedure prior to sequencing, dot hybridizations to the template preparations were performed to determine whether both D L K P E N and G T P E Y L A P E probes recognize the same strand. For clone PSK-K5, hybridization was to opposite strands (not shown). Therefore, this clone was not further characterized.

Each of the three remaining PSK hybridizing fragments was sequenced and found to contain a single open reading frame encoding an amino acid sequence characteristic of the protein-serine kinase family (Fig. 3). All three contain nucleotides encoding the primary target sequence D L K P E N. For PSK-C3, this sequence contains one position of probe mismatch, while all codons in PSK-H1 and PSK-J3 are represented in the probe. PSK-C3 and PSK-H1 both also encode an amino acid sequence very closely related to the secondary target consensus, G T P *G* Y L A P E in PSK-C3 and G T P E Y *I* A P E in PSK-H1. The secondary target sequences contained in PSK-C3 and PSK-H1 match sequences in the G T P E Y L A P E probe at 20 and 23 of the 26 positions, respectively. The PSK-J3 secondary target encodes the rather highly divergent sequence V T L W Y R A P E. A clear hybridization signal was produced despite the fact that 8 of the 26 nucleotide bases were mismatched with the probe. Interestingly, this sequence is present in probable protein-serine kinases encoded by genes that complement cell-cycle mutations in two species of yeast—*CDC28* in *Saccharomyces cerevisiae* (31) and *CDC2* in *Schizosaccharomyces pombe* (32).

The amino acid sequences encoded by the three *Pst* I fragments were aligned for maximum homology with sequences from related protein-serine kinases (Fig. 4). The PSK-J3 homology with the yeast CDC2/CDC28 kinases extends beyond the probe target sequences (Fig. 4*A*). Of the 101 amino acids deduced, 47 are identical to corresponding residues in the CDC2 sequence and 48 show identity to CDC28 residues. Seventy-four amino acids in the PSK-J3 sequence show either identity or a high degree of functional homology with a corresponding CDC amino acid residue. The PSK-C3 clone encodes a peptide that is closely related to a portion of the phosphorylase kinase catalytic γ-subunit (Fig. 4*B*), showing 72% identity with the sequence obtained from a protein isolated from rabbit skeletal muscle. PSK-H1 shows no extensive homology with any of the protein-serine kinases whose amino acid sequence has been reported. The closest relationships are to myosin light-chain kinase (42% sequence identity) and cAMP-dependent protein kinase (41% sequence identity), while the other sequences aligned in Fig. 4*B* have identities ranging from 34% to 38%.

DISCUSSION

Mixed sequence oligonucleotide probes represent a powerful tool for identifying DNA clones that encode a protein whose amino acid sequence has been wholly or partially determined (reviewed in ref. 33). The lower limit of effective probe length for colony hybridization is in the range of 14–17 nucleotides (29, 34), representing five to six consecutive codons. Shorter probes, however, have been successfully used in secondary screening procedures where hybridization is to purified plasmid DNA (34). In this report, I have demonstrated the use of oligonucleotide probe mixtures to identify cDNA clones encoding previously uncharacterized members of a protein family. Sequential screenings with probes directed at two distinct conserved regions within the protein-serine kinase catalytic domain have resulted in the identification of clones from a HeLa cDNA library that appear to encode unique members of this family of enzymes. Given the high proliferative capacity of HeLa cells, some of these encoded proteins may be expected to play a role in regulating cell proliferation. By using the same probes to screen libraries prepared from other cell types or tissues, it is likely that a set of clones encoding protein-serine kinases with different regulatory functions would be identified.

The homology probing approach described here should also be useful for identifying clones encoding previously unknown or poorly characterized members of other protein families; the prerequisite being amino acid sequence information from several members of the family and highly homologous regions of five or more contiguous amino acids within those sequences. For this purpose, short probes recognizing only highly conserved regions are ideal. The use of longer fragments of a previously cloned gene or cDNA under low-stringency hybridization conditions can result in high background signal due to nonspecific annealing of divergent sequences.

Two of the three HeLa putative protein-serine kinase clones that were characterized by partial DNA sequence analysis encode amino acid sequences that show a close

```
PSK-C3:  GluThrArgSerIleMetArgSerLeuLeuGluAlaValSerPheLeuHisProAsnAsnIleValHisArgAspLeuLysProGluAsnIleLeuLeuAspAspAsn
         GAAACCAGGTCCATCATGCGGTCTCTCCTGGAGGCAGTGAGCTTTCTCCATCCCAACAACATTGTGCATCGAGATCTAAAGCCCGAGAATATTCTCCTAGATGACAAT

         MetGlnIleArgLeuSerAspPheGlyPheSerSerHisLeuAspProGlyGluLysLeuArgGluLeuCysGlyThrProGlyTyrLeuAlaProGluIleLeuLysCysSerMet
         ATGCAGATCCGACTTTCAGATTTCGGGTTCTCCTCCCACTTGGACCCTGGCGAGAAGCTTCGAGAGTTGTGTGGGACCCCAGGTTATCTAGCGCCAGAGATCCTTAAATGCTCCATG

         AspGluThrHisProGlyTyrGlyLeuValAspAspLeuTrpAlaCysGlyGluIleLeuPheThrLeuLeuAlaGlySerProProPheTrpHisArgArgGlnIleLeuMetLeu
         GATGAAACCCACCCAGGCTATGGCTTGGTGGACGACCTCTGGGCCTGTGGGGAGATCTTGTTCACACTCCTGGCTGGCTCGCCACCCTTCTGGCACCGGCGGCAGATCCTGATGTTA

         ArgMetIleMetGluGlyGlnTyrGlnPheSerSerProGluTrpAspAspArgSerSerThrValLysAspLeuIleSerArgLeuLeuGln
         CGCATGATCATGGAGGGCCAGTACCAGTTCAGTTCCCCCGAGTGGGATGACCGTTCCAGCACTGTCAAAGACCTGATCTCCAGGCTGCTGCAG

PSK-H1:  LeuGlnMetValLeuAspGlyValArgTyrLeuHisAlaLeuGlyIleThrHisArgAspLeuLysProGluAsnLeuLeuTyrTyrHisProGlyThrAspSerLys
         CTGCAGATGGTCCTGGATGGCGTCCGGTATCTGCATGCACTGGGCATCACACACCGAGACCTCAAACCTGAGAATCTGCTCTACTACCATCCGGGCACTGACTCCAAG

         IleIleIleThrAspPheGlyLeuAlaSerAlaArgLysLysGlyAspAspCysLeuMetLysThrThrCysGlyThrProGluTyrIleAlaProGluValLeuValArgLysPro
         ATCATCATCACCGACTTCGGCCTGGCCAGTGCTCGCAAGAAGGGTGATGACTGCTTGATGAAGACCACCTGTGGCACGCCTGAGTACATTGCCCCAGAAGTCCTGGTCCGCAAGCCA

         TyrThrAsnSerValAspMetTrpAlaLeuGlyValIleAlaTyrIleLeuSerGlyThrMetProPheGluAspAspAsnArgThrArgLeuTyrArgGlnIleLeuArgGly
         TACACCAACTCAGTGGACATGTGGGCGCTGGGCGTCATTGCCTACATCCTACTCAGTGGCACCATGCCGTTTGAGGATGACAACCGTACCCGGCTGTACCGCCAGATCCTCAGGGGC

         LysTyrSerTyrSerGlyGluProTrpProSerValSerAsnLeuAlaLysAspPheIleAspArgLeuLeuThrValAspProGlyAlaArgMetThrAlaLeuGln
         AAGTACAGTTACTCTGGGGAGCCCTGGCCTAGTGTGTCCAACCTGGCCAAGGACTTCATTGACCGCCTGCTGACAGTGGACCCTGGAGCCCGTATGACTGCACTGCAG

PSK-J3:  LysValThrLeuValPheGluHisValAspGlnAspLeuArgThrTyrLeuAspLysAlaProProProGlyLeuProAlaGluThrIleLysAspLeuMetArgGln
         AAGGTAACCCTGGTGTTTGAGCATGTAGACCAGGACCTAAGGACATATCTGGACAAGGCACCCCCACCAGGCTTGCCAGCCGAAACGATCAAGGATCTGATGCGCCAG

         PheLeuArgGlyLeuAspPheLeuHisAlaAsnCysIleValHisArgAspLeuLysProGluAsnIleLeuValThrSerGlyGlyThrValLysLeuAlaAspPheGlyLeuAla
         TTTCTAAGAGGCCTAGATTTCCTTCATGCCAATTGCATCGTTCACCGAGATCTGAAGCCAGAGAACATTCTGGTGACAAGTGGTGGAACAGTCAAGCTGGCTGACTTTGGCCTGGCC

         ArgIleTyrSerTyrGlnMetAlaLeuThrProValValValThrLeuTrpTyrArgAlaProGluValLeuLeuGln
         AGAATCTACAGCTACCAGATGGCACTTACACCCGTGGTTGTTACACTCTGGTACCGAGCTCCCGAAGTTCTTCTGCAG
```

FIG. 3. DNA sequences of hybridizing *Pst* I fragments from clones PSK-C3, PSK-H1, and PSK-J3, and their deduced amino acid sequences. Sequences targeted by probes are underlined. Nucleotide bases within the target sequences having a complementary base in the probe mixture are indicated by asterisks.

A

```
PSK-J3          K V T L V F E H V D Q D L R T Y L D K A - - - P P G L P A E T I K D L M R Q F L R G L D F L H A N C I V
CDC2  (79)      K L Y L V F E F L D M D L K K Y M D R I S E T G A T S L D P R L V Q K F T Y Q L V N G V N F C H S R R I I
CDC28 (84)      K L Y L V F E F L D L D L K R Y M E G I P K D Q P L G A D - - I V K K F M M Q L C K G I A Y C H S H R I L

PSK-J3          H R D L K P E N I L V T S G G T V K L A D F G L A R I Y S Y Q M - A L T P V V V T L W Y R A P E V L L Q
CDC2  (132)     H R D L K P Q N L L I D K E G N L K L A D F G L A R S F G V P L R N Y T H E I V T L W Y R A P E V L L G
CDC28 (135)     H R D L K P Q N L L I N K D G N L K L G D F G L A R A F G V P L R A Y T H E I V T L W Y R A P E V L L G
```

B

```
PSK-C3          E T R S I M R S L L E A V S F L H P N N I V H R D L K P E N I L L D D - - N M Q I R L - S D F G F S S - H L
PhK-γ  (125)    E T R K I M R A L L E V I C A L H K L N I V H R D L K P E N I L L D D - - D M N I K L - T D F G F S C - Q L
PKC-α  (439)    Q A V F Y A A E I S I G L F F L H K R G I I Y R D L K L D N V M L D S E - G H - I K I - A D F G M C K E H M
cGPK   (459)    T T R F Y T A C V V E A F A Y L H S K G I I Y R D L K P E N L I L D H R - G - Y A K L - V D F G F A K - K I
cAPK   (142)    H A R F Y A A Q I V L T F E Y L H S L D L I Y R D L K P E N L L I D Q Q - G - Y I Q V - T D F G F A K - R V
MLCK   (157)    V D T M V F R Q I C D G I L F M H K M R V L H L D L K P E N I L C V N T T G H L V K I - I D F G L A R - R Y
PSK-H1                  L Q M V L D G V R Y L H A L G I T H R D L K P E N L L Y Y H P G T D S K I I I T D F G L A S A R K

PSK-C3          D P G E K L - R E L C G T P G Y L A P E I L K C S M D E T H P G Y G L Y D D L W A C G E I L F T L L A G S P
PhK-γ  (174)    D P G E K L - R E V C G T P S Y L A P E I I E C S M N D N H P G Y G K E V D M W S T G V I M Y T L L A G S P
PKC-α  (489)    M D G V T - T R T F C G T P D Y I A P E I I - - - - - - A Y Q P Y G K S V D W W A Y G V L L Y E M L A G Q P
cGPK   (508)    - G F G K K T W T F C G T P E Y V A P E I I - - - - - - L N K G H D I S A D Y W S L G I L M Y E L L T G S P
cAPK   (191)    K G R - - - T W T L C G T P E Y L A P E I I - - - - - - L S K G Y N K A V D W W A L G V L I Y E M A A G Y P
MLCK   (209)    N P N E K L - K V N F G T P E F L S P E V V - - - - - - N Y D Q I S D K T D M W S L G V I T Y M L L S G L S
PSK-H1          K G D D C L M K T T C G T P E Y I A P E V L - - - - - - V R K P Y T N S V D M W A L G V I A Y I L L S G T M

PSK-C3          P F W N R R Q I L M L R M I M E G Q Y Q F S S P E W D D R S S T V K D L I S R L L Q
PhK-γ  (228)    P F W H R K Q M L M L R M I M S G N Y Q F G S P E W D D Y S D T V K D L V S R F L V V Q P Q K R Y T A E E
PKC-α  (537)    P F D Q E D E D E L F Q S I M E H N V S Y - - - - P K S L S K E A V S I C K G L M T K H P G K R L G C G P
cGPK   (558)    P F S G P D P M K T Y N I I L R G - I D M I E F - P K K I A K N A A N L I K K L C R D N P S E R L G N L K
cAPK   (237)    P F F A D Q P I Q I Y E K I V S G K V R F - - - - P S H F S S D L K D L L R N L L Q V D L T K R F G N L K
MLCK   (257)    P F L G D D D T E T L N N V L S G N W Y F D E E T F E A V S D E A K D F V S N L I Y K E Q G A R M S A A Q
PSK-H1          P F E D D N R T R L Y R Q I L R G K Y S Y S G E P W P S V S N L A K D F I D R L L T V D P G A R M T A L Q
```

FIG. 4. Alignment of amino acid sequences deduced from PSK *Pst* I fragments with sequences from protein-serine kinases. The sequences were aligned by eye for maximum homology and gaps were introduced as indicated by dashes. Numbers in parentheses indicate position from amino termini. Positions of amino acid identity are enclosed within boxes. (*A*) Alignment of PSK-J3 fragment with partial sequences deduced from yeast *CDC2* (32) and *CDC28* (31) genes. (*B*) Alignment of PSK-C3 and PSK-H1 *Pst* I fragments with partial sequences from mammalian protein-serine kinases. Sequences shown are PhK-γ, cGPK, cAPK, and MLCK as described in Fig. 1 legend; PKC-α, protein kinase C (α form) from bovine brain (10). Only identities between PSK-C3 and PhK-γ and between PSK-H1 and MLCK are boxed.

relationship to previously reported sequences. The PSK-C3 sequence shares an identity of 72% with the γ-subunit of phosphorylase kinase isolated from rabbit skeletal muscle (4). This high level of homology suggests a recent evolutionary divergence and possible similarity of function. The PSK-J3 deduced amino acid sequence shows nearly 50% identity to sequences from two yeast protein-serine kinases: CDC2 from *S. pombe* (32) and CDC28 from *S. cerevisiae* (31). The *CDC2* and *CDC28* gene products appear to have equivalent regulatory functions (reviewed in ref. 35). Each is required for progression through the cell cycle at two distinct points; once in G_1 phase prior to initiation of DNA synthesis and again leading into mitosis. The ability of the *CDC28* gene to complement the *CDC2* mutation (36) further demonstrates the functional relationship between these two gene products. A mammalian equivalent of these cell-cycle regulating enzymes has yet to be identified. The putative protein-serine kinase encoded by the PSK-J3 RNA may play a corresponding role in regulation of the mammalian cell cycle.

The amino acid sequence deduced from the PSK-H1 fragment contains the conserved residues characteristic of the protein-serine kinase family as a whole but shows no exceptional homology to any of the protein-serine kinases whose sequence has been determined. The greatest identities, slightly >40%, are with myosin light-chain kinase (6) and cAMP-dependent protein kinase (3). The availability of a cDNA clone provides a means for further characterization of this putative protein-serine kinase.

In addition to the 4 clones described in this report (PSK-C3, PSK-H1, PSK-J3, and PSK-K5), 15 others were recognized by the two oligonucleotide probe mixtures and thus appear to be good candidates for membership in the protein-serine kinase family. These remain to be further characterized by DNA sequence analysis.

The continued acquisition of sequence data for the protein-serine kinases will undoubtedly suggest directions for probe refinement. For example, while this report was in preparation, deduced amino acid sequences for another mammalian protein-serine kinase, protein kinase C, were published (9, 10, 37, 38). Part of the sequence from the α form, isolated from bovine brain (10), is included in the alignment shown in Fig. 4*B*. Each of the reported protein kinase C sequences encode the amino acid sequences D L K L D N and G T P D Y I A P E in positions corresponding to the primary and secondary probe targets. The D L K P E N probe described in Fig. 1 contains complementary bases to the protein kinase C sequences in only 14 or 15 of the 17 positions, depending on the form and species. This level of mismatch makes it uncertain whether protein kinase C clones could have been detected. Redesign of the primary probe mixture to include these codon possibilities would ensure a higher match frequency.

Note Added in Proof. Recently, a homology probing approach has been used to identify another member of the *ras* gene family (39).

Numerous colleagues at the Salk Institute have offered technical advice; Jim Boulter, Paul Gardner, Detlev Jaehner, Chris Saris, and Gerry Weinmaster were especially helpful. I also thank Jim Woodgett and Kathy Gould for comments on the manuscript, Tony Hunter for making available a manuscript prior to publication, and Callie Mack for synthesis of oligonucleotides. A special thanks to Bob Holley for providing laboratory space, support, and encouragement. This work was supported by National Institutes of Health National Research Service Award GM09391 from the National Institute of General Medical Sciences.

1. Krebs, E. G. & Beavo, J. A. (1979) *Annu. Rev. Biochem.* **48**, 923–959.
2. Cohen, P. (1982) *Nature (London)* **296**, 613–620.
3. Shoji, S., Parmelee, D. C., Wade, R. D., Kumar, S., Ericsson, L. H., Walsh, K. A., Neurath, H., Long, G. L., Demaille, J. G., Fischer, E. H. & Titani, K. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 848–851.
4. Reimann, E. M., Titani, K., Ericsson, L. H., Wade, R. D., Fischer, E. H. & Walsh, K. A. (1984) *Biochemistry* **23**, 4185–4192.
5. Takio, K., Wade, R. D., Smith, S. B., Krebs, E. G., Walsh, K. A. & Titani, K. (1984) *Biochemistry* **23**, 4207–4218.
6. Takio, K., Blumenthal, D. K., Edelman, A. M., Walsh, K. A., Krebs, E. G. & Titani, K. (1985) *Biochemistry* **24**, 6028–6037.
7. Downward, J., Yarden, Y., Mayes, E., Scrace, G., Totty, N., Stockwell, P., Ullrich, A., Schlessinger, J. & Waterfield, M. D. (1984) *Nature (London)* **307**, 521–527.
8. Ullrich, A., Bell, J. R., Chen, E. Y., Herrera, R., Petruzelli, L. M., Dull, T. J., Gray, A., Coussens, L., Liao, Y.-C., Tsubokawa, M., Mason, A., Seeburg, P. H., Grunfeld, C., Rosen, O. M. & Ramachandran, J. (1985) *Nature (London)* **313**, 756–761.
9. Ono, Y., Kurokawa, T., Kawahara, K., Nishimura, O., Marumoto, R., Igarashi, K., Sugino, Y., Kikkawa, U., Ogita, K. & Nishizuka, Y. (1986) *FEBS Lett.* **203**, 111–115.
10. Parker, P. J., Coussens, L., Totty, N., Rhee, L., Young, S., Chen, E., Stabel, S., Waterfield, M. D. & Ullrich, A. (1986) *Science* **233**, 853–859.
11. Sefton, B. M. & Hunter, T. (1984) *Adv. Cyclic Nucleotide Protein Phosphorylation Res.* **18**, 195–226.
12. Hunter, T. & Cooper, J. A. (1986) in *The Enzymes*, eds. Boyer, P. D. & Krebs, E. G. (Academic, Orlando, FL), pp. 191–246.
13. Barker, W. C. & Dayhoff, M. O. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 2836–2839.
14. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299.
15. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1408–1412.
16. Okayama, H. & Berg, P. (1983) *Mol. Cell. Biol.* **3**, 280–289.
17. Hanahan, D. (1983) *J. Mol. Biol.* **166**, 557–580.
18. Sinha, N. D., Biernat, J., McManus, J. & Koster, H. (1984) *Nucleic Acids Res.* **12**, 4539–4557.
19. Grunstein, M. & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 3961–3965.
20. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
21. Jay, E., Bambara, R., Padmanabhan, R. & Wu, R. (1974) *Nucleic Acids Res.* **1**, 331–354.
22. Wood, W. I., Gitschier, J., Lasky, L. A. & Lawn, R. M. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1585–1588.
23. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513–1523.
24. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
25. Meinkoth, J. & Wahl, G. (1984) *Anal. Biochem.* **138**, 267–284.
26. Sanger, F. & Coulson, A. R. (1975) *J. Mol. Biol.* **94**, 441–448.
27. Norrander, J., Kempe, T. & Messing, J. (1983) *Gene* **26**, 101–106.
28. Kafatos, F. C., Jones, C. W. & Efstratiadis, A. (1979) *Nucleic Acids Res.* **7**, 1541–1552.
29. Lathe, R. (1985) *J. Mol. Biol.* **183**, 1–12.
30. Takeya, T., Feldman, R. A. & Hanafusa, H. (1982) *J. Virol.* **44**, 1–11.
31. Lorincz, A. T. & Reed, S. I. (1984) *Nature (London)* **307**, 183–185.
32. Hindley, J. & Phear, G. A. (1984) *Gene* **31**, 129–134.
33. Itakura, K., Rossi, J. J. & Wallace, R. B. (1984) *Annu. Rev. Biochem.* **53**, 323–356.
34. Suggs, S. V., Wallace, R. B., Hirose, T., Kawashima, E. H. & Itakura, K. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 6613–6617.
35. Nurse, P. (1985) *Trends Genet.* **1**, 51–55.
36. Beach, D., Durkacz, B. & Nurse, P. (1982) *Nature (London)* **300**, 706–709.
37. Coussens, L., Parker, P. J., Rhee, L., Yang-Feng, T. L., Chen, E., Waterfield, M. D., Francke, U. & Ullrich, A. (1986) *Science* **233**, 859–866.
38. Knopf, J. L., Lee, M.-H., Sultzman, L. A., Kriz, R. W., Loomis, C. R., Hewick, R. M. & Bell, R. M. (1986) *Cell* **46**, 491–502.
39. Chardin, P. & Tavitian, A. (1986) *EMBO J.* **5**, 2203–2208.